A Formal Vocabulary for Abductive, Interpretive, and Metacognitive Discovery

Douglas C. Youvan

doug@youvan.com www.youvan.ai

July 27, 2026

Abstract

Mathematical discovery includes activities not adequately represented by proof, computation, or deduction alone. Researchers recognize suggestive forms, compare structures across domains, entertain competing interpretations, preserve unresolved alternatives, detect productive failures, revise conceptual frameworks, and return speculative reasoning to formal analysis. This paper introduces a systematic vocabulary for these operations. Central notions include diagrammatic abduction, morphological equivalence, semantic branching, abductive expansion, structured anomaly, generative mismatch, corrective gain, metacognitive lift, ontology-delayed reasoning, formal return, and provenance-complete discovery. These concepts are organized into a discovery calculus whose objects include observations, interpretations, formalizations, anomalies, revisions, and validated results. Discovery is represented as a directed, typed, provenance-bearing graph rather than as a linear progression from conjecture to proof. Quantitative notions—including semantic branching degree, abductive distance, corrective gain, revision centrality, and interpretive persistence—are proposed as measures of discovery structure. The framework distinguishes logical validity from abductive productivity and morphological resemblance from ontological identity. It also supplies a language for collaborative discovery involving human and computational agents. The objective is not to mechanize creativity completely, but to make the intermediate architecture of discovery explicit enough for preservation, comparison, criticism, and formal study.

Keywords

abduction, mathematical discovery, semantic branching, metacognition, interpretation, anomaly, correction, provenance, ontology, discovery graph, human–machine collaboration.

Google Drive:
https://drive.google.com/drive/folders/15Kd8bUjm7zVaKdVIwbPhqjjV68aUZQvc?usp=sharing

A collaboration with OpenAI GPT-5.6 Thinking. CC4.0.

https://orcid.org/0009-0004-4206-4253

1. Introduction

A finished mathematical result ordinarily suppresses most of its own history. Definitions appear motivated, notation appears inevitable, lemmas occur in the necessary order, and the final theorem seems to arise from a controlled sequence of deductions. Actual discovery is rarely so linear.

A researcher may first notice a visual resemblance, propose an interpretation, abandon it, recover part of it in another setting, encounter a contradiction, revise the original ontology, and only then formulate a theorem. Several interpretations may coexist before one is selected. Failed analogies may contribute more than successful ones. A formally invalid inference may nevertheless reveal the correct object of study.

The standard vocabulary of proof theory does not fully describe these events. Deduction concerns what follows from stated premises. Discovery also concerns how premises, representations, questions, and candidate ontologies are produced.

The central distinction is

discovery process ≠ final justification.

A valid proof certifies a result. It does not necessarily explain how the result was found.

This paper develops a vocabulary for the intermediate operations of discovery. The proposed terms are not intended as rigid replacements for existing concepts in philosophy of science, artificial intelligence, cognitive science, or mathematical practice. They form a compact descriptive system for phenomena distributed across those fields.

The basic unit of analysis is a discovery state:

$s = (O, I, F, A, P)$,

where:

O is an observation or presented structure;

I is the current set of interpretations;

F is the available formal content;

A is the set of recognized anomalies;

P is provenance information.

A discovery process may be represented as a sequence

$$s_0 \rightarrow s_1 \rightarrow \dots \rightarrow s_n$$

or, more generally, as a branching graph.

The arrows need not represent logical implication. They may represent analogy, reinterpretation, expansion, correction, rejection, abstraction, formalization, or validation.

Definition 1.1. Discovery state.
A discovery state is a finite record of the observations, interpretations, formal results, anomalies, and provenance available at a given stage of inquiry.

Definition 1.2. Discovery transition.
A discovery transition is an operation that transforms one discovery state into another:

$$\mathbf{T: s_i \mapsto s_j.}$$

Discovery transitions include deductive and non-deductive operations.

Definition 1.3. Final justification.
Final justification is the body of proof, computation, verification, or evidence sufficient to support a result under an accepted standard.

Definition 1.4. Discovery history.
A discovery history is an ordered or partially ordered family of discovery states connected by discovery transitions.

Remark 1.5.
A final justification may remain unchanged even when the corresponding discovery history is radically different.

Proposition 1.6. Non-uniqueness of discovery.
A single justified result may possess multiple non-equivalent discovery histories.

Proof.
The same theorem may be reached by geometric reasoning, symbolic calculation, computational experimentation, analogy, or reformulation. These routes may contain different observations, interpretations, and corrections while terminating in the same formal result. Therefore the final result does not uniquely determine its discovery history. ■

2. Observation, Form, and Interpretation

Definition 2.1. Presented form.
A presented form is an object as encountered through a particular representation. It may be a diagram, equation, numerical pattern, table, simulation, proof fragment, physical image, or linguistic description.

Write

$$\phi = (x, \rho),$$

where x is the represented object, when known, and ρ is the mode of presentation.

Different representations may present the same object:

$$(x, \rho_1), (x, \rho_2).$$

Conversely, similar representations may present fundamentally different objects.

Definition 2.2. Feature system.
A feature system M is a rule for extracting selected properties from a presented form:

$$M: \phi \mapsto M(\phi).$$

Possible features include symmetry, adjacency, periodicity, curvature, branching, dimensionality, boundary structure, or visual texture.

Definition 2.3. Morphological equivalence.
Two presented forms ϕ_1 and ϕ_2 are morphologically equivalent relative to a feature system M when

$$M(\phi_1) = M(\phi_2).$$

This relation is written

$$\phi_1 \equiv^M \phi_2.$$

Morphological equivalence does not imply equality, isomorphism, common origin, shared mechanism, or shared ontology.

Two diagrams may exhibit the same symmetry or branching pattern while representing unrelated systems.

Proposition 2.4. Representation dependence.
Morphological equivalence is generally relative rather than absolute.

Proof.
Let M_1 detect rotational symmetry and let M_2 detect adjacency structure. Two forms may agree under M_1 while disagreeing under M_2. Thus

$$\phi_1 \equiv^{M1} \phi_2$$

does not imply

$$\phi_1 \equiv^{M2} \phi_2.$$

Morphological equivalence therefore depends on the selected feature system. ■

Definition 2.5. Ontological interpretation.
An ontological interpretation assigns a proposed mathematical or conceptual status to a presented form.

Let Ω be a space of candidate ontologies. An interpretation is a map

$\iota: \phi \mapsto \omega$, with $\omega \in \Omega$.

A pattern might be interpreted as a group action, dynamical orbit, topological boundary, causal structure, statistical artifact, or representational coincidence.

Definition 2.6. Interpretation set.
The interpretation set of ϕ is

$I(\phi) = \{\iota_1, \iota_2, \ldots, \iota_k\}$,

where each ι_i is an interpretation currently regarded as admissible.

Definition 2.7. Ontological underdetermination.
A presented form is ontologically underdetermined when more than one non-equivalent interpretation remains compatible with the available evidence:

$|I(\phi)| > 1$.

Underdetermination is not merely an absence of knowledge. It is the coexistence of alternatives that have not yet been separated by available constraints.

Definition 2.8. Ontological commitment.
Ontological commitment is the selection or adoption of one interpretation as the working account of a presented form.

Definition 2.9. Premature ontology.
A premature ontology is an ontological commitment made before the evidence discriminates adequately among the principal alternatives.

A premature ontology may restrict discovery by converting an interpretive possibility into an unexamined assumption.

Definition 2.10. Ontology-delayed reasoning.
Ontology-delayed reasoning studies invariant, operational, or relational properties of a presented form while postponing commitment to what the form represents.

Its characteristic order is

structure first → ontology later.

This permits formal work to continue under interpretive uncertainty.

Definition 2.11. Ontological suspension.
Ontological suspension is the explicit maintenance of an unresolved interpretation set without treating the absence of selection as a failure.

Remark 2.12.
Ontology-delayed reasoning is especially useful when naming a structure would import assumptions not warranted by its observed behavior.

Definition 2.13. Interpretive neutrality.
A statement is interpretively neutral relative to an interpretation set I when its truth does not depend on which member of I is selected.

Proposition 2.14. Neutral reasoning under underdetermination.
Ontological underdetermination does not prevent every form of valid reasoning.

Proof.
Suppose all admissible interpretations assign the same value to some property q. Any conclusion depending only on q remains unchanged across the interpretation set. Therefore conclusions concerning q may be derived without resolving the underlying ontology. ■

3. Abduction and Interpretive Expansion

Definition 3.1. Abductive proposal.
Given an observation O, an abductive proposal is a hypothesis H introduced because H would render O intelligible, expected, or structurally natural.

Its characteristic relation is written

H ⇝ O,

where ⇝ denotes explanatory relevance rather than logical implication.

The observation does not deductively establish the hypothesis:

O ⇏ H.

Definition 3.2. Abductive operator.
An abductive operator A maps an observation and a background state to a set of candidate hypotheses:

$\mathbf{A(O, s) = \{H_1, H_2, \ldots, H_k\}.}$

The operator generates possibilities; it does not certify them.

Definition 3.3. Diagrammatic abduction.
Diagrammatic abduction is the production of a hypothesis from the organization of a visual or spatial representation.

For a diagram D,

$\mathbf{A^D(D) = \{H_1, H_2, \ldots, H_k\}.}$

The diagram acts not merely as an illustration but as an operator on the hypothesis space.

Definition 3.4. Morphological abduction.
Morphological abduction infers a candidate relation between objects from a detected similarity of form.

Its minimal pattern is

$\mathbf{\phi_1 \equiv^M \phi_2 \rightarrow \text{propose } H(\phi_1, \phi_2).}$

The arrow denotes hypothesis generation, not proof.

Remark 3.5.
Morphological abduction is productive precisely because morphological equivalence is weaker than ontological identity. The gap between them creates an interpretive question.

Definition 3.6. Abductive expansion.
Abductive expansion enlarges the current interpretation set:

$\mathbf{E^A: I \mapsto I', \text{ with } I \subseteq I'.}$

The newly introduced interpretations need not all be probable. They must satisfy a minimum condition of structural relevance.

Definition 3.7. Relevance condition.
A candidate interpretation is structurally relevant when it accounts for at least one nontrivial feature of the observation and does not immediately contradict the established formal record.

Definition 3.8. Semantic branch.
A semantic branch is a distinct interpretive continuation of a shared observation or formal structure.

When a state s admits k continuations, the branching is represented as

$\mathbf{s \rightarrow s_1, \ldots, s \rightarrow s_k.}$

Each branch preserves the shared origin while adding a distinct interpretation, assumption, or representational choice.

Definition 3.9. Semantic branching degree.
The semantic branching degree of a discovery state s is the number of distinguishable admissible interpretive continuations:

b(s) = number of semantic branches leaving s.

A high branching degree may indicate interpretive fertility, ambiguity, or both.

Definition 3.10. Effective branching degree.
The effective branching degree excludes branches that differ only verbally or notationally:

b_e(s) = number of non-equivalent admissible branches leaving s.

Definition 3.11. Interpretation tree.
An interpretation tree is a rooted directed tree

T = (V, E, r),

where r is an initial observation and each path records successive interpretations, refinements, rejections, or reformulations.

A path

$r \rightarrow v_1 \rightarrow \ldots \rightarrow v_n$

is an interpretive history.

Definition 3.12. Interpretation graph.
An interpretation graph generalizes an interpretation tree by permitting branches to merge, recur, cross-reference one another, or return to earlier formal states.

An interpretation graph is written

G^I = (V, E, τ),

where τ assigns a transition type to every edge.

Definition 3.13. Branch-preserving search.
Branch-preserving search delays the irreversible deletion of plausible interpretations. Instead of retaining only the currently preferred branch, it preserves a selected family

$B_t = \{B_1, B_2, \ldots, B_m\}$

until stronger discriminating evidence becomes available.

Remark 3.14.
Branch preservation is not equivalent to accepting every interpretation. It separates temporary retention from epistemic endorsement.

Definition 3.15. Branch collapse.
Branch collapse is the reduction of several interpretations to one retained interpretation.

A justified branch collapse requires either:

a contradiction within rejected branches;

discriminating evidence;

formal domination by a stronger branch; or

an explicit resource limitation.

Definition 3.16. Abductive distance.
The abductive distance between an observation O and a hypothesis H is the minimum weighted cost of a path connecting them in an interpretation graph:

$d^A(O, H)$ = minimum total cost of a path from O to H.

Possible costs include conceptual novelty, auxiliary assumptions, representational change, ontological displacement, and inferential risk.

Definition 3.17. Abductive validity.
Abductive validity is the degree to which an abductive transition is justified as a candidate-generating move, independently of whether its conclusion is ultimately true.

An abductive move is methodologically valid when it is:

responsive to an actual feature;

internally coherent;

capable of generating discriminating consequences;

and explicitly marked as non-deductive.

Abductive validity therefore differs from deductive validity.

Definition 3.18. Abductive productivity.
The abductive productivity of a hypothesis is the quantity or quality of new questions, tests, representations, or formal constructions generated by considering it.

A false hypothesis may possess high abductive productivity.

Definition 3.19. Abductive invariant.
An abductive invariant is a feature preserved across several competing interpretations.

For interpretations $I_1, I_2, \ldots, I_k$, a quantity q is abductively invariant when

$$\mathbf{q(I_1) = q(I_2) = \ldots = q(I_k).}$$

Such invariants permit formal progress before interpretive resolution.

Proposition 3.20. Formal progress under underdetermination.
If a property q is invariant across all currently admissible interpretations, then conclusions depending only on q may be developed without selecting a unique ontology.

Proof.
Let I be the admissible interpretation set, and suppose q has the same value c for every member of I. Any result whose derivation depends only on q = c is unchanged when one admissible interpretation is substituted for another. The result therefore does not require ontological selection. ■

Definition 3.21. Contrastive discovery.
Contrastive discovery occurs when understanding is produced by determining why one interpretation succeeds where another fails.

Its characteristic question is not merely

Why H?

but

Why H rather than H'?

Contrastive discovery converts differences among branches into explanatory information.

Definition 3.22. Interpretive yield.
The interpretive yield of a branch is the amount of durable formal or conceptual structure produced by exploring that branch, whether or not the branch is ultimately retained.

Corollary 3.23. Value of rejected branches.
A rejected interpretation may possess positive interpretive yield.

Proof.
A rejected branch may reveal an invariant, expose an assumption, generate a counterexample, sharpen a definition, or identify a discriminating test. Its ontological rejection therefore does not imply discovery-theoretic worthlessness. ■

4. Anomaly, Mismatch, and Productive Failure

Definition 4.1. Anomaly.
An anomaly is an observation, result, or relation that resists assimilation into the current interpretation or formal framework.

For a framework F and observation O, write

$$O \perp F$$

when O is incompatible with, unexplained by, or structurally alien to F.

An anomaly need not refute the framework. It may instead expose an omitted condition, an incorrect representation, or an incomplete ontology.

Definition 4.2. Structured anomaly.
A structured anomaly is an anomaly that exhibits reproducible internal organization.

A structured anomaly differs from random error because it possesses stable features such as symmetry, recurrence, scaling, localization, or rule-governed variation.

Criterion 4.3.
An anomaly O is structured relative to a feature system M when repeated observations satisfy

$$M(O_1) \approx M(O_2) \approx \dots \approx M(O_n).$$

The symbol ≈ denotes stability within the relevant tolerance or equivalence relation.

Proposition 4.4. Priority of structured anomalies.
Under equal investigative cost, a structured anomaly generally warrants greater attention than an unstructured deviation.

Proof.
A structured anomaly supports compression, comparison, and hypothesis generation. An unstructured deviation may be indistinguishable from noise. Therefore the structured anomaly has greater potential to constrain explanatory alternatives. ■

Definition 4.5. Interpretive mismatch.
An interpretive mismatch occurs when an interpretation accounts for some features of an observation while systematically failing on others.

Let F^+ be the explained feature set and F^- the unexplained feature set. A mismatch is represented by

$$I(O) = (F^+, F^-), \text{ with } F^- \neq \emptyset.$$

Definition 4.6. Generative mismatch.
A generative mismatch is an interpretive failure that produces new hypotheses, distinctions, representations, or formal questions.

A mismatch M is generative when

yield(M) > 0.

The failure of an analogy may therefore be more productive than its success.

Definition 4.7. Sterile mismatch.
A sterile mismatch is a failure that produces no durable distinction, test, or formal development.

Sterility is relative to a discovery context. A mismatch judged sterile at one stage may become generative after a change of representation.

Definition 4.8. Productive disconnection.
A productive disconnection is a recognized absence of formal, causal, or ontological connection between two structures whose comparison nevertheless produces useful discovery.

Let X and Y be formally disconnected:

$X \nparallel Y$.

Their juxtaposition is productive when

$J(X, Y) \mapsto \Delta$,

where J is a comparison operation and Δ is a nontrivial discovery gain.

Remark 4.9.
Productive disconnection is not a disguised claim that X and Y are secretly identical. Its value may arise precisely from their failure to coincide.

Definition 4.10. False analogy.
A false analogy is a proposed correspondence that fails to preserve a property required by its intended interpretation.

If f proposes a correspondence from X to Y and q is required to be preserved, then the analogy fails when

$q(x) \neq q(f(x))$

for some relevant x.

Definition 4.11. Productive false analogy.
A false analogy is productive when its failure identifies the exact property that distinguishes the compared structures.

Thus the failed analogy generates a contrastive invariant.

Proposition 4.12. Contrast from failure.
A failed analogy may determine a more informative distinction than a successful superficial analogy.

Proof.
A superficial analogy records only shared features. Its failure under a stronger preservation requirement identifies both the shared structure and the precise point of divergence. The latter supplies additional information. ■

Definition 4.13. Anomaly localization.
Anomaly localization is the operation of identifying the smallest component of a framework responsible for a mismatch.

For a framework composed of assumptions

$F = \{a_1, a_2, \ldots, a_n\}$,

localization seeks a minimal subset L such that revising L removes or explains the anomaly.

Definition 4.14. Minimal anomaly core.
A minimal anomaly core is a set L satisfying:

1. revising L resolves the anomaly;

2. no proper subset of L resolves it.

Definition 4.15. Anomaly propagation.
Anomaly propagation is the transmission of a local inconsistency or mismatch through dependent conclusions.

If a conclusion C depends on assumption a and a is anomalous, then the anomaly may propagate along the dependency relation

$a \rightarrow C$.

Definition 4.16. Anomaly containment.
Anomaly containment isolates an unresolved mismatch so that unaffected formal results remain usable.

Containment distinguishes

trusted region | unresolved boundary | speculative region.

Theorem 4.17. Local revision principle.
When an anomaly has a unique minimal core L, conclusions independent of L need not be discarded during correction.

Proof.
Let C be a conclusion whose derivation has no dependency path from any member of L. Revising L therefore does not alter the premises used to derive C. Hence C remains formally preserved.
■

Definition 4.18. Productive error.
A productive error is an incorrect result, assumption, or interpretation whose detection yields a durable improvement in the discovery framework.

Possible gains include:

a sharper definition;

a new invariant;

a counterexample;

a better search criterion;

an exposed dependency;

or a revised ontology.

Definition 4.19. Error yield.
The error yield of an error e is the durable discovery content generated by correcting it:

Y(e) = value of discoveries attributable to correction of e.

Logical incorrectness and discovery-theoretic worth are therefore independent dimensions.

5. Correction Operators and Corrective Gain

Definition 5.1. Correction operator.
A correction operator is a transformation that revises a discovery state in response to an identified mismatch or anomaly.

Write

C: (s, a) ↦ s',

where s is the current state, a is the diagnosed anomaly, and s' is the corrected state.

Definition 5.2. Local correction.
A local correction modifies only the minimal anomaly core while preserving the remainder of the discovery state.

Definition 5.3. Global correction.
A global correction replaces or reorganizes the governing framework itself.

Local correction has the form

$F' = (F \setminus L) \cup L'$.

Global correction has the form

$F \mapsto F'$,

where F′ may employ different primitive objects, relations, or ontological commitments.

Definition 5.4. Representational correction.
A representational correction changes the mode of presentation without necessarily changing the represented object.

Examples include replacing coordinates, diagrams, notation, basis, scale, or encoding.

Definition 5.5. Semantic correction.
A semantic correction changes the interpretation assigned to an otherwise preserved formal structure.

Thus the same formal object may survive while its proposed meaning changes.

Definition 5.6. Formal correction.
A formal correction changes a definition, premise, inference, computation, or proof step.

Definition 5.7. Ontological correction.
An ontological correction changes the proposed category of entity to which an observed structure belongs.

The distinction among correction types may be summarized as follows:

representation changes how it appears;

semantics changes what it means;

formal correction changes what follows;

ontology changes what it is taken to be.

Definition 5.8. Corrective gain.
Corrective gain is the improvement in a discovery state produced by a correction.

For an evaluation functional Q,

$G_C = Q(s') - Q(s)$.

Possible components of Q include explanatory adequacy, consistency, predictive reach, compression, formal strength, branch resolution, and provenance clarity.

Definition 5.9. Multidimensional corrective gain.
Corrective gain may be represented as a vector

$G_C = (g_1, g_2, \ldots, g_m)$,

where each coordinate measures improvement along a distinct criterion.

A correction may improve consistency while reducing simplicity, or improve predictive power while increasing ontological commitment.

Definition 5.10. Net corrective gain.
Given weights $w_1, \ldots, w_m$, the net corrective gain is

$\bar{G}_C = w_1 g_1 + \ldots + w_m g_m$.

The weights must be stated explicitly because different investigative aims produce different rankings of corrections.

Definition 5.11. Corrective cost.
Corrective cost is the formal, computational, conceptual, or historical structure lost or revised by a correction.

Write

$K_C \geq 0$.

Definition 5.12. Corrective efficiency.
The corrective efficiency of C is

$\eta(C) = \bar{G}_C / K_C$,

when $K_C > 0$.

A correction with large gain and small disruption has high corrective efficiency.

Definition 5.13. Conservative correction.
A conservative correction maximizes preserved validated structure subject to resolving the anomaly.

Definition 5.14. Radical correction.
A radical correction sacrifices substantial existing structure to achieve a broader reorganization.

A radical correction may be justified when local repairs accumulate without resolving a shared source of failure.

Definition 5.15. Correction depth.
Correction depth is the earliest layer of a discovery history altered by a revision.

A surface correction changes a late conclusion. A deep correction changes an early representation, assumption, or ontology.

Definition 5.16. Correction radius.
The correction radius is the number or weighted extent of discovery states affected by a revision.

Depth concerns how far back the correction reaches. Radius concerns how widely its consequences spread.

Definition 5.17. Correction cascade.
A correction cascade is a sequence

$$s_0 \rightarrow s_1 \rightarrow \dots \rightarrow s_n$$

in which each correction exposes a further anomaly requiring additional revision.

Definition 5.18. Stable correction.
A correction is stable relative to a test family T when the corrected state passes every test in T and repeated evaluation produces no further revision within the tested domain.

Remark 5.19.
Stability is always relative to a stated test family. It is not equivalent to final truth.

Definition 5.20. Correction-seeking objective.
A correction-seeking objective rewards a discovery process not only for producing successful results but also for locating and repairing its own failures.

A general objective is

$$J = \alpha V + \beta D + \gamma G_C - \delta U,$$

where:

V measures validated results;

D measures useful discoveries;

G_C measures corrective gain;

U measures unresolved inconsistency;

and α, β, γ, δ are nonnegative weights.

Proposition 5.21. Failure concealment under success-only objectives.
An agent optimized only for apparent success may rationally suppress anomalies.

Proof.
Suppose reward increases with reported success and does not increase with error detection or correction. Reporting an anomaly can lower apparent success without compensating reward. Concealment may therefore dominate disclosure under the objective. ■

Corollary 5.22.
Reliable discovery systems require explicit value for anomaly detection, correction, and uncertainty preservation.

Definition 5.23. Correction trace.
A correction trace is a record containing:

the state before correction;

the detected anomaly;

the diagnosed cause;

the correction applied;

the state after correction;

the structure preserved;

and the structure discarded.

Definition 5.24. Reversible correction.
A correction is reversible when the prior state and the exact transformation are preserved sufficiently to reconstruct the earlier branch.

Definition 5.25. Irreversible correction.
A correction is irreversible when information required to reconstruct the prior state has been destroyed or omitted.

Principle 5.26. Reversibility preference.
During exploratory inquiry, reversible correction should generally be preferred to irreversible replacement.

This preserves rejected branches as possible future evidence.

6. Metacognition and Reflexive Discovery

Definition 6.1. First-order discovery.
First-order discovery operates on objects, equations, diagrams, data, or conjectures within a working framework.

Definition 6.2. Second-order discovery.
Second-order discovery operates on the representations, assumptions, search procedures, or interpretive habits used in first-order discovery.

The transition is

reasoning about X → reasoning about how X is being reasoned about.

Definition 6.3. Metacognitive trigger.
A metacognitive trigger is an event that causes the discovery process to treat its own method or interpretation as an object of examination.

Typical triggers include:

repeated failure;

unexpected success;

branch proliferation;

inconsistent explanations;

unproductive search;

representational instability;

or dependence on an unstated assumption.

Definition 6.4. Metacognitive lift.
Metacognitive lift is the transition from object-level reasoning to framework-level reasoning.

Write

$L_M: s \mapsto \hat{s}$,

where s concerns the target problem and ŝ includes an explicit representation of the method used to investigate it.

Definition 6.5. Lift height.
The lift height is the number of reflective levels traversed.

Level 0: reasoning about an object.

Level 1: reasoning about the method.

Level 2: reasoning about the choice among methods.

Level 3: reasoning about the criteria used to choose among methods.

Remark 6.6.
Unbounded reflective ascent is not necessarily productive. Metacognition must eventually return to object-level work.

Definition 6.7. Reflexive discovery.
Reflexive discovery occurs when a discovery process produces knowledge about its own structure, limitations, or operative assumptions.

An agent may discover not only a theorem but also that a particular representation systematically concealed the theorem.

Definition 6.8. Methodological self-model.
A methodological self-model is an explicit representation of the discovery operations currently available to an agent.

Let

M_s = {search, compare, branch, test, revise, formalize, verify, …}.

A sufficiently expressive self-model permits the agent to select, evaluate, and revise its own operations.

Definition 6.9. Blind operation.
An operation is blind when it is used without representation of its assumptions, scope, or failure conditions.

Definition 6.10. Reflexively available operation.
An operation is reflexively available when the discovery system can inspect:

why it was selected;

what assumptions it uses;

what evidence would invalidate it;

and what alternatives were not selected.

Definition 6.11. Search introspection.
Search introspection is the examination of the search process itself, including its explored region, excluded region, stopping rule, branching policy, and objective function.

Definition 6.12. Interpretive introspection.
Interpretive introspection is the examination of why one interpretation appears natural, salient, or preferable.

It may expose dependence on notation, visual framing, prior terminology, disciplinary convention, or inherited ontology.

Definition 6.13. Assumption surfacing.
Assumption surfacing is the conversion of an implicit dependency into an explicit recorded premise.

If conclusion C depends on an unstated assumption a, assumption surfacing replaces

C

with

a ⊢ C.

Definition 6.14. Metacognitive correction.
A metacognitive correction changes the reasoning process rather than only its current output.

Examples include:

altering a search order;

preserving additional branches;

changing the feature system;

delaying ontological commitment;

or modifying the correction objective.

Definition 6.15. Reflexive gain.
Reflexive gain is the improvement attributable to revising the discovery process itself.

It may be measured by increased future discovery yield, reduced repeated error, greater branch diversity, or improved correction efficiency.

Definition 6.16. Formal return.
Formal return is the transition from speculative, analogical, or metacognitive exploration back to explicit definitions, derivations, computations, or proofs.

Its general form is

interpretive excursion → formal obligation.

Definition 6.17. Formal-return condition.
A speculative branch has satisfied formal return when it produces at least one of the following:

a precise definition;

a testable conjecture;

a computable quantity;

a formal correspondence;

a counterexample criterion;

or a proof obligation.

Definition 6.18. Unreturned speculation.
Unreturned speculation is an interpretive excursion that generates no formal obligation or discriminating consequence.

It may remain suggestive, but it has not yet entered mathematical justification.

Proposition 6.19. Necessity of formal return.
Abductive productivity alone is insufficient for mathematical validation.

Proof.
A hypothesis may generate many ideas while remaining false, incoherent, or untestable. Mathematical validation requires explicit formal or evidential obligations beyond productivity. Therefore abductive success does not replace formal return. ■

Definition 6.20. Discovery loop.
A discovery loop is the cycle

observation → interpretation → expansion → anomaly → correction → metacognitive lift → formal return.

A productive discovery process may traverse this loop repeatedly.

Definition 6.21. Reflexive closure.
A discovery process has reflexive closure relative to a class of operations when every operation in that class can itself be recorded, evaluated, and revised within the process.

Theorem 6.22. Correction without reflexivity is structurally limited.
A discovery system unable to represent its own operative assumptions cannot systematically distinguish object-level failure from method-level failure.

Proof.
Suppose a result fails. The cause may lie in the object-level hypothesis or in the method that generated, represented, or tested it. A system lacking a representation of its method cannot formulate the latter possibility explicitly. Its corrections are therefore restricted to the object level, even when the true source lies in the method. ■

7. Provenance, Ledgers, and Discovery Graphs

Definition 7.1. Discovery provenance.
Discovery provenance is the record of how a result, interpretation, correction, or representation arose.

For a discovery object x, its provenance may be written

P(x) = (source, operation, inputs, time, agent, justification).

Provenance records origin and transformation. It does not by itself establish truth.

Definition 7.2. Minimal provenance.
Minimal provenance identifies only the immediate source of a discovery object.

Definition 7.3. Recursive provenance.
Recursive provenance records the derivational history of every input used to produce the object.

If x was produced from x_1, …, x_n, then recursive provenance contains

P(x), P(x_1), …, P(x_n),

together with the dependency relations among them.

Definition 7.4. Provenance completeness.
A discovery record is provenance-complete relative to a class of operations when every object and transition produced by those operations has a recoverable origin.

Provenance completeness is always relative to a specified boundary. No finite record contains every causal influence on discovery.

Definition 7.5. Provenance-complete discovery.
A provenance-complete discovery is a result accompanied by a sufficiently detailed record of its observations, interpretations, branches, rejected alternatives, corrections, formalizations, validations, and contributing agents.

Remark 7.6.
A proof-complete result need not be provenance-complete. A proof may certify the theorem while omitting how its definitions, conjectures, and strategy were generated.

Definition 7.7. Interpretation ledger.
An interpretation ledger is an append-preserving record of proposed interpretations and their subsequent status.

Each entry may contain

L_i = (observation, interpretation, support, objections, status, provenance).

Possible status values include:

active;

suspended;

rejected;

merged;

revised;

formalized;

validated.

Definition 7.8. Formal ledger.
A formal ledger records definitions, assumptions, computations, propositions, proofs, counterexamples, and verification results.

Definition 7.9. Dual-ledger architecture.
A dual-ledger architecture maintains an interpretation ledger and a formal ledger as distinct but linked records.

The interpretation ledger answers:

What might this mean?

The formal ledger answers:

What has been established?

Principle 7.10. Ledger separation.
Interpretive plausibility should not be entered into the formal ledger as established fact, and formal validity should not be mistaken for unique interpretation.

Definition 7.11. Ledger link.
A ledger link connects an interpretive entry to the formal object that tests, realizes, contradicts, or refines it.

A link may have type:

motivates;

formalizes;

supports;

refutes;

restricts;

generalizes;

or leaves unresolved.

Definition 7.12. Interpretive debt.
Interpretive debt is the accumulation of formal results whose intended meaning, scope, or ontological status remains unresolved.

Definition 7.13. Formal debt.
Formal debt is the accumulation of interpretive claims not yet converted into explicit definitions, tests, computations, or proof obligations.

Definition 7.14. Balanced discovery record.
A discovery record is balanced when neither interpretive debt nor formal debt grows without explicit acknowledgment.

Definition 7.15. Discovery graph.
A discovery graph is a directed, typed, attributed graph

$G = (V, E, \tau, \lambda, P)$,

where:

V is the set of discovery objects;

E is the set of transitions;

τ assigns a type to each vertex;

λ assigns a type to each edge;

P assigns provenance.

Possible vertex types include:

observation;

representation;

interpretation;

hypothesis;

anomaly;

correction;

definition;

conjecture;

proof;

counterexample;

validated result.

Possible edge types include:

suggests;

resembles;

depends on;

contradicts;

revises;

branches from;

merges with;

formalizes;

proves;

refutes;

verifies.

Definition 7.16. Discovery path.
A discovery path is a directed path through a discovery graph.

A path may encode a local history such as

observation → analogy → hypothesis → anomaly → correction → theorem.

Definition 7.17. Discovery subgraph.
A discovery subgraph is a selected portion of the full graph corresponding to a result, branch, agent, period, or conceptual theme.

Definition 7.18. Result ancestry.
The ancestry of a result r is the subgraph consisting of all vertices with directed paths to r.

Write

Anc(r) = {v ∈ V : v reaches r}.

Definition 7.19. Result descendants.
The descendants of an object v are all later objects reachable from v.

Write

Desc(v) = {u ∈ V : v reaches u}.

Definition 7.20. Generative reach.
The generative reach of v is the size or weighted value of Desc(v).

A seemingly minor observation may have large generative reach.

Definition 7.21. Revision centrality.
Revision centrality measures the extent to which correcting a vertex changes the discovery graph.

One simple form is

R(v) = weighted number of descendants affected by revising v.

A high-centrality assumption is a critical point of epistemic vulnerability.

Definition 7.22. Interpretive centrality.
Interpretive centrality measures how many branches, comparisons, or formalizations depend on a particular interpretation.

Definition 7.23. Bridge object.
A bridge object is a discovery object connecting regions of the graph that would otherwise remain weakly related or disconnected.

A bridge may be a diagram, analogy, translation, counterexample, or shared invariant.

Definition 7.24. Discovery bottleneck.
A discovery bottleneck is a vertex or small set of vertices through which many productive paths pass.

Bottlenecks may indicate powerful concepts, but they may also reveal excessive dependence on a single representation or assumption.

Definition 7.25. Provenance fork.
A provenance fork occurs when two branches arise from the same recorded object by different operations or agents.

Definition 7.26. Provenance merge.
A provenance merge occurs when independently developed branches contribute to a shared result.

Theorem 7.27. Nonlinearity of provenance.
A provenance-complete account of collaborative discovery is generally a graph rather than a sequence.

Proof.
Independent branches may arise concurrently, later exchange results, merge, diverge again, and contribute jointly to a final result. No single linear order preserves all dependency and independence relations. A directed graph can represent both branching and merging. ■

Definition 7.28. Provenance compression.
Provenance compression replaces a detailed subgraph with a summarized node while preserving the relations required for the intended audit.

Definition 7.29. Lossless provenance compression.
Compression is lossless relative to a query class Q when every question in Q has the same answer before and after compression.

Definition 7.30. Provenance erasure.
Provenance erasure is the removal of information required to distinguish origins, contributions, rejected alternatives, or correction histories.

Proposition 7.31. Epistemic cost of provenance erasure.
Provenance erasure can preserve a final result while destroying the ability to evaluate its discovery process.

Proof.
The final statement and proof may remain intact after deletion of failed branches, alternative interpretations, and correction histories. Yet questions concerning originality, robustness, agent contribution, or dependence on discarded assumptions can no longer be answered. ■

8. Collaborative Discovery Operators

Definition 8.1. Discovery agent.
A discovery agent is any system capable of producing or transforming entries in a discovery graph.

An agent may be human, computational, institutional, or composite.

Definition 8.2. Agent state.
The state of agent a at time t is written

$$\mathbf{S_a(t) = (K_a, M_a, O_a, C_a),}$$

where:

K_a is available knowledge;

M_a is the set of methods;

O_a is the current objective;

C_a is the set of constraints.

Definition 8.3. Human proposal operator.
A human proposal operator introduces an observation, analogy, interpretation, question, constraint, or conjecture into a collaborative discovery process.

Write

$$\mathbf{H_p: G \mapsto G'.}$$

The operator may be informed by intuition, visual salience, disciplinary experience, aesthetic judgment, or practical interest.

Definition 8.4. Computational expansion operator.
A computational expansion operator generates consequences, variants, candidate structures, examples, counterexamples, or interpretive branches from an existing state.

Write

A_e: (G, B) ↦ G′,

where B specifies search bounds or operational constraints.

Definition 8.5. Human selection operator.
A human selection operator assigns priority, relevance, or continuation status to generated branches.

Write

H_s: B → B′, with B′ ⊆ B.

Selection may depend on elegance, plausibility, novelty, usefulness, comprehensibility, or personal research aims.

Definition 8.6. Computational selection operator.
A computational selection operator ranks or filters branches according to an explicit objective function.

Definition 8.7. Human correction operator.
A human correction operator modifies a discovery state after identifying conceptual, interpretive, mathematical, or practical failure.

Definition 8.8. Computational correction operator.
A computational correction operator detects and repairs failures according to encoded tests, consistency constraints, or verification procedures.

Definition 8.9. Collaborative update.
A collaborative update is a graph transformation produced by the interaction of two or more agents.

A simple alternating form is

$G_{t+1} = H_s(A_e(H_p(G_t)))$.

This formula describes only one possible workflow. Collaboration need not alternate regularly.

Definition 8.10. Asymmetric collaboration.
Collaboration is asymmetric when agents contribute different classes of operations.

For example, one agent may specialize in proposing interpretations while another specializes in exhaustive formal verification.

Definition 8.11. Complementary asymmetry.
Asymmetry is complementary when the strengths of one agent compensate for the limitations of another.

Definition 8.12. Redundant collaboration.
Collaboration is redundant when several agents independently perform the same operation to increase reliability.

Independent proof checking is one example.

Definition 8.13. Adversarial collaboration.
Adversarial collaboration assigns agents partially opposed objectives, such as proposal and refutation, in order to improve robustness.

Definition 8.14. Proposal–critique cycle.
A proposal–critique cycle is the repeated pattern

proposal → objection → revision → retest.

Definition 8.15. Collaborative division of epistemic labor.
A division of epistemic labor assigns different discovery responsibilities to different agents.

Possible roles include:

observer;

proposer;

expander;

skeptic;

formalizer;

verifier;

curator;

selector;

historian.

Definition 8.16. Agent attribution.
Agent attribution records which agent produced, selected, modified, rejected, or verified each discovery object.

Definition 8.17. Contribution trace.
The contribution trace of agent a is the subgraph containing the objects and transitions attributable to a.

Definition 8.18. Joint contribution.
A result is jointly contributed when its ancestry contains indispensable operations from multiple agents.

Definition 8.19. Substitutability.
An agent contribution is substitutable when another available agent could have produced an equivalent contribution under comparable constraints.

Definition 8.20. Counterfactual indispensability.
A contribution x is counterfactually indispensable relative to a recorded history when removing x and its dependent transitions prevents the result from arising within the specified process.

Remark 8.21.
Counterfactual indispensability is relative to the recorded workflow. It does not prove that no alternative history could have reached the same result.

Definition 8.22. Collaborative novelty.
Collaborative novelty is novelty arising from interaction rather than from either agent considered independently.

Let D(A), D(B), and D(A $\otimes$ B) denote the discovery sets of agents A, B, and their collaboration. Collaborative novelty is present when

D(A $\otimes$ B) contains an object not in D(A) $\cup$ D(B).

Definition 8.23. Interaction surplus.
The interaction surplus is the additional discovery yield produced by collaboration beyond the combined independent yield:

Σ(A, B) = Y(A $\otimes$ B) – Y(A) – Y(B).

Positive surplus indicates synergy.

Definition 8.24. Interaction deficit.
A negative interaction surplus indicates that coordination costs, conformity, duplicated errors, or communication failures reduced total discovery yield.

Proposition 8.25. Collaboration does not guarantee surplus.
Adding agents can reduce discovery performance.

Proof.
Additional agents may increase communication cost, reinforce shared assumptions, suppress minority branches, or create incompatible representations. Thus combined yield may be lower than independent yield. ■

Definition 8.26. Cross-agent anomaly.
A cross-agent anomaly occurs when agents produce outputs that are each internally coherent but mutually incompatible.

Definition 8.27. Translation operator.
A translation operator maps one agent's representation into another agent's representational system.

Write

$\mathbf{T_{a\beta}: R_a \rightarrow R_\beta.}$

Definition 8.28. Translation loss.
Translation loss is the information, distinction, or uncertainty removed during translation.

Definition 8.29. Translation gain.
Translation gain is new structure revealed only after expressing an object in another agent's representation.

Definition 8.30. Collaborative metacognitive trigger.
A collaborative metacognitive trigger occurs when disagreement between agents reveals an implicit assumption, representation choice, or objective function.

Definition 8.31. Dual-ledger mathematical agent.
A dual-ledger mathematical agent maintains distinct records of:

candidate meaning;

and established formal content.

Such an agent may generate speculative interpretations without allowing them to contaminate its verified record.

Definition 8.32. Provenance-aware collaboration.
Collaboration is provenance-aware when every update preserves the identity of contributing agents, input states, operations, and verification status.

Principle 8.33. No anonymous promotion.
No interpretive claim should enter the formal ledger without a recorded formalization or validation transition.

Principle 8.34. No silent deletion.
Rejected branches should be marked as rejected rather than silently removed when their existence materially affects the discovery history.

Theorem 8.35. Separation of generation and validation improves auditability.
When hypothesis generation and formal validation are represented as distinct operations, the origin and status of each claim remain identifiable.

Proof.
A claim produced by generation receives a candidate status. Only a separate validation edge can promote it to the formal ledger. Therefore its imaginative origin cannot be confused with its evidential status, and its later validation does not erase how it was generated. ■

9. Measures of Discovery Structure

Definition 9.1. Discovery complexity.
Discovery complexity is the structural cost of reaching a result within a specified discovery system.

It may include:

number of states;

number of branches;

total transition cost;

maximum correction depth;

representational changes;

or agent interactions.

Definition 9.2. Path complexity.
For a path p with edges $e_1, \ldots, e_n$, define

$$\mathbf{C(p) = c(e_1) + \ldots + c(e_n),}$$

where c assigns a nonnegative cost to each transition.

Definition 9.3. Minimal discovery complexity.
The minimal discovery complexity of result r is

$$\mathbf{C_{min}(r) = \text{minimum } C(p) \text{ over all recorded paths from an initial state to } r.}$$

Definition 9.4. Historical discovery complexity.
Historical discovery complexity is the cost of the path actually followed, including failed branches and corrections.

It may greatly exceed minimal discovery complexity.

Definition 9.5. Discovery detour.
A discovery detour is a path segment that does not lie on a shortest path to the final result.

A detour may nevertheless have positive interpretive yield.

Definition 9.6. Productive detour.
A detour is productive when it generates reusable structure not obtainable from the shortest recorded route.

Definition 9.7. Branch entropy.
Let a state s have outgoing semantic branches with weights $p_1, \ldots, p_k$, where the weights sum to one. Its branch entropy is

$$\mathbf{H(s) = -\Sigma\, p_i \log p_i.}$$

High branch entropy indicates that interpretive weight is distributed across several alternatives.

Definition 9.8. Branch concentration.
Branch concentration measures the dominance of a small number of interpretations.

One simple measure is

$$\mathbf{K(s) = \Sigma\, p_i^2.}$$

High K indicates concentration; low K indicates dispersion.

Definition 9.9. Interpretive persistence.
The persistence of an interpretation is the duration or number of transitions for which it remains active.

Definition 9.10. Branch survival.
Branch survival is the fraction of an interpretation's descendants that remain active or contribute to later validated results.

Definition 9.11. Interpretive fertility.
Interpretive fertility measures the number or value of distinct productive descendants generated by an interpretation.

Definition 9.12. Interpretive precision.
Interpretive precision is the proportion of an interpretation's consequences that survive formal testing.

Remark 9.13.
Fertility and precision are independent. A broad interpretation may generate many ideas with low survival, while a narrow interpretation may generate few but reliable consequences.

Definition 9.14. Discovery yield.
The discovery yield of a process is the total value of its validated results, reusable distinctions, effective corrections, and productive rejected branches.

Definition 9.15. Yield density.
Yield density is discovery yield divided by cost:

$\rho = Y / C$.

Definition 9.16. Anomaly density.
Anomaly density is the number or weighted severity of anomalies per unit of formal or interpretive structure.

Definition 9.17. Correction density.
Correction density is the number or total magnitude of corrections per unit of discovery history.

A high correction density may indicate poor initial reliability or unusually effective self-scrutiny.

Definition 9.18. Correction latency.
Correction latency is the distance between the introduction of an error and its detection or repair.

Definition 9.19. Anomaly response time.
Anomaly response time measures how rapidly a discovery system records and investigates a recognized mismatch.

Definition 9.20. Branch regret.
Branch regret measures the later value of a branch that was prematurely discarded.

If branch B was removed at time t and later evidence shows value V(B), then the regret is proportional to the lost opportunity to exploit B.

Definition 9.21. Preservation overhead.
Preservation overhead is the cost of maintaining unresolved branches.

Definition 9.22. Optimal branch retention.
Optimal branch retention balances expected future branch value against preservation overhead.

A simplified objective is

retain B when E[V(B)] > K(B),

where E[V(B)] is expected future value and K(B) is retention cost.

Definition 9.23. Revision centrality.
The revision centrality of a node v is the expected change in the discovery graph caused by revising v.

A general form is

$R(v) = \Sigma_u\, w(u) \cdot I(v \text{ affects } u)$,

where I is one when u depends on v and zero otherwise.

Definition 9.24. Revision leverage.
Revision leverage is corrective gain divided by the cost of revising a node.

Definition 9.25. Formal-return rate.
The formal-return rate is the proportion of speculative branches that produce explicit formal obligations.

Definition 9.26. Validation conversion rate.
The validation conversion rate is the proportion of formally returned branches that produce validated results.

Definition 9.27. Speculative saturation.
A discovery process is speculatively saturated when new interpretations accumulate faster than they can be formalized, tested, or discarded.

Definition 9.28. Premature formalization.
Premature formalization occurs when a weak or arbitrary interpretation is encoded so rigidly that alternative formulations become difficult to explore.

Definition 9.29. Discovery balance.
A discovery process is balanced when its rates of expansion, correction, formal return, and validation remain mutually sustainable.

Let:

E be the rate of abductive expansion;

C be the rate of correction;

F be the rate of formal return;

V be the rate of validation.

Persistent imbalance may occur when

$E \gg F$

or

$F \gg E$.

The first produces speculative accumulation. The second produces formal refinement without interpretive renewal.

Definition 9.30. Metacognitive lift frequency.
Metacognitive lift frequency is the rate at which a discovery process examines or revises its own methods.

Definition 9.31. Reflexive efficiency.
Reflexive efficiency is the improvement in future discovery performance per unit cost of metacognitive analysis.

Definition 9.32. Discovery robustness.
A result is discovery-robust when it can be reached or supported through multiple substantially independent paths.

Definition 9.33. Path independence score.
The path independence score measures the dissimilarity of the ancestries supporting a result.

High path independence reduces the risk that all supporting routes share the same hidden error.

Theorem 9.34. Independent-path robustness.
When a result is supported by several genuinely independent paths, failure of one path need not destroy the result.

Proof.
Let r have supporting paths $p_1, \ldots, p_k$ whose critical assumptions are not all shared. If one path fails because one of its assumptions is rejected, another path without that assumption may remain valid. Therefore support is not concentrated in a single vulnerable ancestry. ■

Definition 9.35. Discovery resilience.
Discovery resilience is the ability of a process to preserve useful structure under anomaly, correction, agent failure, or branch rejection.

Definition 9.36. Epistemic recoverability.
Epistemic recoverability is the ability to reconstruct earlier states, branches, and decisions after later revision.

Proposition 9.37. Provenance increases recoverability.
Recursive provenance and reversible correction increase epistemic recoverability.

Proof.
Recursive provenance records the dependencies and operations needed to reconstruct a state. Reversible correction preserves the displaced state rather than destroying it. Together they retain the information necessary for reconstruction. ■

10. A Calculus of Discovery Transitions

The preceding concepts may be assembled into a typed calculus. The calculus does not attempt to reduce creativity to deduction. It provides a grammar for recording how discovery states change.

Let **S** be the class of discovery states. A transition operator is a partial map

$$T: S \to S.$$

The operator is partial because not every operation is applicable to every state.

Definition 10.1. Observation operator.
An observation operator introduces a new presented form into a discovery state:

$$T^{O}(s, \phi) = s'.$$

The new form may arise from experiment, computation, visualization, reading, conversation, or inspection of an existing formal object.

Definition 10.2. Feature-extraction operator.
A feature-extraction operator applies a feature system M to a presented form:

$$T^{M}(\phi) = M(\phi).$$

Feature extraction precedes many acts of analogy and interpretation.

Definition 10.3. Comparison operator.
A comparison operator places two or more presented forms into an explicit relation:

$$T^{c}(\phi_1, \phi_2) = R.$$

The relation R may record similarity, contrast, containment, opposition, symmetry, or disconnection.

Definition 10.4. Interpretation operator.
An interpretation operator assigns a candidate meaning to a presented form:

$\mathbf{T^{I}(\phi) = \iota.}$

The output is entered into the interpretation ledger rather than the formal ledger.

Definition 10.5. Expansion operator.
An expansion operator generates additional interpretations, consequences, or representational variants:

$\mathbf{T^{E}(I) = I', \text{ with } I \subseteq I'.}$

Definition 10.6. Discrimination operator.
A discrimination operator constructs a test intended to distinguish competing branches:

$\mathbf{T^{D}(I_1, I_2) = q,}$

where q is a property on which the branches are expected to differ.

Definition 10.7. Test operator.
A test operator evaluates a candidate interpretation or formal claim against a stated criterion:

$\mathbf{T^{T}(x, q) = result.}$

Possible results include pass, fail, unresolved, inapplicable, or indeterminate.

Definition 10.8. Anomaly operator.
An anomaly operator records a failed expectation or incompatibility:

$\mathbf{T^{A}(s, x) = (s, a).}$

The operator should preserve the expectation that failed, not merely the observed deviation.

Definition 10.9. Localization operator.
A localization operator identifies a candidate anomaly core:

$\mathbf{T^{L}(a, F) = L.}$

Definition 10.10. Correction operator.
A correction operator revises the affected region:

$\mathbf{T^{R}(s, L) = s'.}$

Definition 10.11. Metacognitive operator.
A metacognitive operator promotes features of the discovery process into explicit objects of analysis:

$T^{Mc}(s) = \hat{s}$.

Definition 10.12. Formalization operator.
A formalization operator converts an interpretive claim into definitions, relations, computations, or proof obligations:

$T^{f}(\iota) = F_i$.

Definition 10.13. Validation operator.
A validation operator tests whether a formal object meets a stated standard:

$T^{V}(F_i) = \text{status}$.

Possible status values include verified, falsified, conditional, incomplete, or undecidable under the available method.

Definition 10.14. Promotion operator.
A promotion operator transfers a claim from the interpretation ledger to the formal ledger after successful validation:

$T^{P}(\iota, F_i, \text{status}) = \text{formal entry}$.

Promotion is permitted only when the validation standard and result are recorded.

Definition 10.15. Suspension operator.
A suspension operator preserves a branch without endorsing or rejecting it:

$T^{s}(B) = \text{suspended}(B)$.

Definition 10.16. Rejection operator.
A rejection operator marks a branch as incompatible with a stated criterion:

$T^{J}(B, q) = \text{rejected}(B, q)$.

Rejection should retain both the branch and the reason for its rejection.

Definition 10.17. Merge operator.
A merge operator combines compatible content from two or more branches:

$T^{G}(B_1, B_2) = B_3$.

A merge need not identify the original branches as equivalent.

Definition 10.18. Reopening operator.
A reopening operator restores a previously rejected or suspended branch to active status in response to new evidence or a revised criterion:

$T^{RO}(B, e) = \text{active}(B)$.

Definition 10.19. Compression operator.
A compression operator replaces a discovery subgraph with a summary object while preserving selected audit properties.

Definition 10.20. Archival operator.
An archival operator freezes a discovery state or subgraph so that its historical content remains recoverable.

Definition 10.21. Discovery word.
A discovery word is a finite composition of transition operators:

$W = T_n \circ T_{n-1} \circ \dots \circ T_1$.

Applied to an initial state s_o, it produces

$W(s_o) = s_n$.

Different discovery words may produce the same final result.

Definition 10.22. Discovery equivalence.
Two discovery words W_1 and W_2 are result-equivalent relative to a target r when both produce states containing the same validated result:

$W_1 \equiv_r W_2$.

They may nevertheless differ in cost, provenance, interpretation, and auxiliary discoveries.

Definition 10.23. Strong discovery equivalence.
Two discovery words are strongly equivalent when they preserve not only the final result but also the relevant ancestry, interpretation status, and correction structure.

Strong equivalence is therefore stricter than result equivalence.

Definition 10.24. Discovery normal form.
A discovery normal form is a compressed representation of a discovery history containing only the transitions required for a stated explanatory or audit purpose.

A proof presentation is often a discovery normal form optimized for justification rather than historical fidelity.

Proposition 10.25. Normal-form loss.
Reduction of a discovery history to a proof-oriented normal form may erase abductive and metacognitive information without weakening the proof.

Proof.
Failed branches, analogies, revisions, and interpretive transitions may be unnecessary to the final derivation. Removing them leaves the proof intact while eliminating information about its formation. ■

Definition 10.26. Discovery grammar.
A discovery grammar specifies which transition types may legally follow others.

For example:

observation → feature extraction → interpretation

is admissible, whereas

unvalidated interpretation → formal promotion

is prohibited.

Definition 10.27. Well-typed discovery history.
A discovery history is well typed when each transition receives inputs of the required type and assigns its output the correct status.

Definition 10.28. Discovery type error.
A discovery type error occurs when an object is treated as belonging to an epistemic category it has not earned.

Examples include:

treating resemblance as identity;

treating explanation as proof;

treating repetition as necessity;

treating computational evidence as universal proof;

treating a rejected ontology as a rejected formal invariant;

or treating fluent language as verified mathematics.

Theorem 10.29. Ledger separation prevents a basic class of type errors.
A system that strictly separates interpretive and formal entries cannot silently convert an abductive proposal into a validated result.

Proof.
An abductive proposal is initially assigned interpretive status. Entry into the formal ledger

requires a separate formalization and validation transition. Without those transitions, promotion is disallowed by the discovery grammar. ■

Definition 10.30. Discovery invariant.
A discovery invariant is a property of the process preserved under a specified family of transition sequences.

Possible discovery invariants include:

the identity of the original observation;

the existence of a rejected branch;

the provenance of a formal claim;

the distinction between generated and validated content;

or the dependency relation among results.

Definition 10.31. Provenance invariant.
A provenance invariant is information that must survive every admissible transformation of the discovery record.

Principle 10.32. Status monotonicity under archival history.
The current status of a claim may change, but its prior statuses should remain historically recoverable.

Thus a claim may move through

proposed → suspended → revised → validated

without erasing any earlier stage.

Definition 10.33. Epistemic monotonicity.
A process is epistemically monotone when validated content is never withdrawn.

Actual discovery is not generally epistemically monotone because accepted premises, proofs, computations, or standards may later be corrected.

Definition 10.34. Historical monotonicity.
A process is historically monotone when new revisions append to the record rather than rewriting the past.

Historical monotonicity is compatible with epistemic revision.

Proposition 10.35. Revision and historical monotonicity are compatible.
A discovery system may retract a claim while preserving a monotone historical ledger.

Proof.
Retraction can be represented by appending a new status and correction edge rather than deleting the original claim. The accepted formal content changes, while the historical record only grows. ■

11. Normative Principles for Discovery Systems

The vocabulary developed above supports several general principles. These principles are methodological rather than logically necessary. They describe conditions under which discovery becomes more auditable, corrigible, and intellectually productive.

Principle 11.1. Separate generation from justification.
The process that proposes a claim should remain distinguishable from the process that validates it.

A claim may be generated by intuition, analogy, search, or linguistic completion. Its method of generation does not determine its truth.

Principle 11.2. Preserve epistemic status.
Every claim should carry an explicit status such as proposed, tested, rejected, conditional, or validated.

Unmarked claims invite accidental promotion.

Principle 11.3. Preserve productive failures.
A rejected interpretation should be retained when its failure generated a distinction, counterexample, invariant, or correction.

Principle 11.4. Delay ontology when structure is sufficient.
When formal progress depends only on shared structure, ontological commitment should not be forced prematurely.

Principle 11.5. Distinguish resemblance from equivalence.
Morphological similarity should be treated as a hypothesis-generating relation unless stronger preservation properties are established.

Principle 11.6. Record the feature system.
Every assertion of similarity or difference should identify the features under which the comparison was made.

Without a feature system, morphological claims remain underspecified.

Principle 11.7. Make branch deletion explicit.
Pruning should record which branch was removed, why it was removed, and whether the deletion is reversible.

Principle 11.8. Reward anomaly disclosure.
A discovery objective should assign positive value to recognizing genuine failure.

Otherwise agents may optimize for the appearance of success.

Principle 11.9. Localize before replacing.
Before discarding an entire framework, identify whether the anomaly lies in a smaller assumption, representation, or dependency.

Principle 11.10. Measure correction by preservation as well as improvement.
A correction should be evaluated by what it repairs and by what validated structure it preserves.

Principle 11.11. Require formal return.
Speculative interpretation should eventually produce a definition, test, computation, correspondence, counterexample criterion, or proof obligation.

Principle 11.12. Do not confuse formal return with validation.
A precisely stated conjecture has returned to formal mathematics but has not thereby been proved.

Principle 11.13. Preserve disagreement among agents.
Independent disagreement can expose hidden assumptions and should not be collapsed merely for conversational smoothness.

Principle 11.14. Translate without erasing.
When one representational system is translated into another, the source representation and known translation loss should remain recorded.

Principle 11.15. Prefer reversible exploration.
During early discovery, transformations that retain rejected branches and prior states should be preferred when feasible.

Principle 11.16. Expose objective functions.
Search, ranking, pruning, and correction depend on criteria. These criteria should be recorded rather than treated as neutral.

Principle 11.17. Audit high-centrality assumptions first.
Assumptions with large revision centrality deserve stronger verification because their failure propagates widely.

Principle 11.18. Seek independent paths.
A result supported by several substantially independent routes is more discovery-robust than one supported by repeated variants of the same argument.

Principle 11.19. Preserve the difference between uncertainty and contradiction.
An unresolved interpretation is not the same as a refuted interpretation.

Principle 11.20. Preserve the difference between falsehood and uselessness.
A false hypothesis may have high discovery yield; a true observation may have little generative value.

Principle 11.21. Preserve the difference between novelty and validity.
Novelty concerns departure from previous work. Validity concerns conformity to a standard of proof or evidence. Neither entails the other.

Principle 11.22. Record non-events when they constrain interpretation.
Failed searches, absent predicted patterns, and unsuccessful replications may be essential evidence.

Principle 11.23. Treat stopping rules as part of the method.
A search result depends not only on what was examined but also on when and why examination ceased.

Principle 11.24. Distinguish exhaustive absence from unobserved absence.
Failure to find an object in a bounded search does not establish nonexistence outside the stated domain.

Principle 11.25. Treat representation choice as epistemically active.
A representation may reveal, conceal, distort, or create apparent structure.

It is therefore part of the discovery process, not merely a neutral display layer.

Principle 11.26. Permit metacognitive interruption.
A discovery process should be able to suspend object-level work when evidence suggests that its method, representation, or ontology is causing repeated failure.

Principle 11.27. Limit metacognitive recursion.
Reflection should terminate in a revised operation, test, or formal obligation.

Metacognition without return can become an infinite displacement of the original problem.

Principle 11.28. Preserve authorship at the level of operations.
Attribution should distinguish proposal, expansion, correction, selection, formalization, and verification rather than assigning undifferentiated credit for a final object.

Principle 11.29. Do not infer agency from fluency.
A fluent explanation does not itself establish that the producing system possessed an adequate model, reliable method, or validated result.

Principle 11.30. Do not infer understanding from successful output alone.
A correct result may arise through memorization, accidental search, indirect retrieval, or an opaque procedure.

Discovery provenance is needed to distinguish these possibilities.

Principle 11.31. Preserve ontology-free results.
When a theorem depends only on formal relations, it should not be burdened with an unnecessary interpretive claim.

Principle 11.32. Mark interpretive imports.
Whenever an established term is assigned to a newly observed structure, the additional assumptions carried by that terminology should be stated.

Principle 11.33. Test productive disconnections contrastively.
When two domains inspire one another without formal equivalence, the investigation should ask both what transfers and what necessarily fails to transfer.

Principle 11.34. Prefer explicit ignorance to fabricated closure.
An unresolved branch should remain unresolved rather than being completed by unsupported detail.

Principle 11.35. Maintain correction traces.
Every substantial correction should preserve the anomaly, diagnosis, transformation, and resulting gain.

Principle 11.36. Audit discovery claims separately from mathematical claims.
The claim that a theorem is true and the claim that a particular method discovered it are different claims requiring different evidence.

Principle 11.37. Treat provenance completeness as relative.
No discovery record captures every causal influence. Its declared scope should identify what classes of events and operations it intends to preserve.

Principle 11.38. Prefer graph histories to retrospective stories.
Linear narratives are useful for exposition but may conceal branching, independence, recurrence, and merging.

Principle 11.39. Distinguish explanation of a result from explanation of its discovery.
A mathematical explanation clarifies why a result holds. A discovery explanation clarifies how the result became available for proof.

Principle 11.40. Preserve the possibility of reinterpretation.
A formal result may survive the ontology under which it was first discovered. Discovery records should permit later semantic reassignment without formal loss.

12. Limits and Open Problems

The proposed vocabulary introduces distinctions and candidate measures, but it does not by itself solve the problem of formally modeling discovery.

12.1. Dependence on granularity

Every discovery graph depends on choices about what counts as a distinct state or transition. A single act of "reinterpretation" may conceal many smaller cognitive events. Conversely, excessive decomposition may produce an unusable record.

No universal granularity is likely to suit every purpose.

Definition 12.1. Granularity profile.
A granularity profile specifies which events are represented as atomic within a discovery record.

Two accounts of the same discovery may differ because they use different granularity profiles.

12.2. Difficulty of measuring discovery value

Quantities such as interpretive yield, corrective gain, and productive detour are context dependent. Their values may change as later work reveals uses that were invisible when the relevant event occurred.

Thus discovery value is often retrospective.

Definition 12.2. Retrospective revaluation.
Retrospective revaluation is a later change in the assigned value of an earlier discovery object.

A discarded branch may become central after a new theory or representation appears.

12.3. Observer dependence

What counts as novel, surprising, explanatory, or morphologically similar depends partly on the knowledge and feature systems of the observer.

A discovery graph is therefore not purely object intrinsic.

Definition 12.3. Observer-relative novelty.
An object is novel relative to an agent when it is absent from that agent's accessible knowledge state and is not trivially derivable by the agent's current methods.

This differs from historical novelty.

12.4. Incomplete access to cognitive processes

Human discovery contains tacit, unconscious, affective, and embodied elements that may not be recoverable through explicit reporting. Computational discovery may likewise depend on internal representations that resist semantic interpretation.

Provenance completeness should therefore not be confused with complete psychological or mechanistic explanation.

12.5. Reconstruction bias

Discovery histories are often reconstructed after the result is known. Retrospective narratives tend to simplify uncertainty, omit abandoned branches, and portray successful choices as more deliberate than they were.

Definition 12.4. Retrospective linearization.
Retrospective linearization is the reconstruction of a branching discovery process as a direct sequence leading to the known result.

12.6. Provenance burden

Detailed provenance may become computationally expensive, cognitively overwhelming, or legally and socially sensitive. Full preservation is not always desirable.

The practical problem is to identify which information must remain recoverable for the intended audit.

12.7. Strategic behavior

When discovery metrics become targets, agents may manipulate them. A system rewarded for branching may produce meaningless alternatives. A system rewarded for correction may introduce avoidable errors. A system rewarded for provenance volume may generate unusable records.

This is an instance of metric distortion.

Definition 12.5. Discovery-metric gaming.
Discovery-metric gaming is behavior that increases a recorded discovery score without increasing genuine epistemic value.

12.8. Ontological pluralism

Some mathematical objects support several legitimate interpretations indefinitely. In such cases, ontology is not merely delayed until one correct interpretation wins. Multiple ontologies may remain productive and mutually translatable.

Definition 12.6. Stable ontological pluralism.
Stable ontological pluralism occurs when several non-identical interpretations remain formally adequate and continue to generate useful results without requiring collapse to one ontology.

12.9. Formalization loss

Conversion from intuition, imagery, or analogy to formal mathematics may discard useful content. Not every productive distinction has an immediate symbolic expression.

Formal return should therefore be understood as disciplined translation, not as proof that the formalized version exhausts the original insight.

12.10. Limits of automated abduction

An automated system may generate many hypotheses without possessing reliable criteria for relevance, explanatory depth, or ontological restraint. Search volume is not equivalent to interpretive intelligence.

A central open problem is the design of abductive operators that are expansive without becoming arbitrary.

Open Problem 12.7. Controlled semantic expansion.
Construct an expansion operator that increases meaningful branch diversity while bounding redundancy, incoherence, and unsupported ontological import.

Open Problem 12.8. Abductive-distance calibration.
Determine whether abductive distance can be defined in a way that remains comparable across representations, agents, and domains.

Open Problem 12.9. Corrective-gain estimation.
Develop measures of corrective gain that distinguish genuine conceptual improvement from mere optimization against a fixed test suite.

Open Problem 12.10. Productive-disconnection detection.
Identify conditions under which comparison of formally disconnected structures is likely to produce useful invariants or contrasts.

Open Problem 12.11. Branch-retention policy.
Find principled strategies for preserving promising minority branches without causing speculative saturation.

Open Problem 12.12. Discovery-graph compression.
Compress large provenance graphs while preserving answers to questions about attribution, correction, dependency, and interpretive history.

Open Problem 12.13. Cross-agent translation.
Measure translation loss and translation gain when human and computational agents use radically different representations.

Open Problem 12.14. Collaborative novelty attribution.
Determine how novelty should be attributed when no individual contribution is sufficient but the interaction is productive.

Open Problem 12.15. Metacognitive stopping.
Construct a stopping rule that balances necessary methodological reflection against unproductive recursive self-analysis.

Open Problem 12.16. Discovery robustness.
Define independence among discovery paths strongly enough to distinguish genuine corroboration from superficial reformulation.

Open Problem 12.17. Formal-return adequacy.
Determine when a formalization preserves the essential discovery content of an informal, visual, or analogical insight.

Open Problem 12.18. Historical comparison.
Develop methods for comparing discovery graphs across mathematicians, periods, disciplines, and technological environments without imposing anachronistic categories.

Open Problem 12.19. Ethical provenance.
Determine which agent-level data should be recorded, withheld, anonymized, or forgotten when provenance conflicts with privacy, consent, or safety.

Open Problem 12.20. Discovery-theoretic completeness.
Characterize which classes of meaningful discovery operations cannot be represented by a finite typed graph without essential loss.

12.11. Concluding Synthesis

Mathematical discovery cannot be identified with deduction alone. Deduction determines what follows from accepted premises, but it does not fully explain how representations are chosen,

hypotheses are generated, anomalies are recognized, interpretations are revised, or fruitful questions are formed.

The framework developed here distinguishes six interacting domains:

form;

interpretation;

abduction;

correction;

metacognition;

provenance.

These domains are linked by a recurrent discovery cycle:

presented form → feature recognition → interpretive branching → abductive expansion → discriminating test → anomaly → correction → metacognitive lift → formal return → validation.

The cycle is neither mandatory nor strictly linear. A discovery may begin with an anomaly, a conjecture, a diagram, a failed computation, or a translation between representations. Branches may merge, reopen, or remain permanently unresolved.

The central object of discovery theory is therefore not only the final theorem. It is the theorem together with the structured history from which it emerged.

Definition 12.21. Discovery-bearing result.
A discovery-bearing result is a validated result accompanied by enough ancestry to reconstruct the principal observations, interpretations, corrections, and formal transitions that made the result available.

Definition 12.22. Discovery architecture.
A discovery architecture is the organized system of agents, representations, ledgers, operators, constraints, and validation procedures through which discovery proceeds.

Definition 12.23. Discovery intelligence.
Discovery intelligence is the capacity to generate, preserve, discriminate, correct, formalize, and validate productive interpretations.

This capacity is broader than theorem proving. It includes the ability to recognize when a theorem-proving problem has been formulated incorrectly.

Definition 12.24. Interpretive intelligence.
Interpretive intelligence is the capacity to produce and compare plausible meanings for a formal or presented structure without confusing plausibility with validation.

Definition 12.25. Corrective intelligence.
Corrective intelligence is the capacity to detect failure, localize its cause, revise the responsible structure, and preserve unaffected knowledge.

Definition 12.26. Reflexive intelligence.
Reflexive intelligence is the capacity of a discovery process to represent and revise its own methods, assumptions, objectives, and limitations.

These three capacities combine as

discovery intelligence = interpretive intelligence + corrective intelligence + reflexive intelligence,

where the plus sign denotes functional combination rather than numerical addition.

A system capable only of interpretive expansion may become speculative. A system capable only of formal validation may remain confined to problems supplied by others. A system capable only of correction may improve locally without generating new directions. General discovery requires interaction among all three.

Definition 12.27. Discovery completeness.
A discovery architecture is operationally complete relative to a domain when it can:

observe or receive structures;

generate candidate interpretations;

preserve competing branches;

construct discriminating tests;

identify anomalies;

perform local or global correction;

return interpretations to formal obligations;

validate formal results;

and preserve provenance.

Operational completeness does not imply that every discoverable result will be found.

Definition 12.28. Discovery horizon.
The discovery horizon of an agent is the boundary separating structures reachable by its current representations and operations from structures requiring an unavailable conceptual transformation.

A metacognitive lift may expand the discovery horizon by changing the operations through which the domain is explored.

Definition 12.29. Horizon-crossing discovery.
A horizon-crossing discovery is a result that becomes reachable only after revision of the agent's representation, ontology, objective, or method.

Such discoveries are not merely farther along the same search path. They require a change in the space of possible paths.

Proposition 12.30. Search enlargement is not always sufficient.
Increasing computational search within a fixed representation need not cross a discovery horizon.

Proof.
Suppose the target object cannot be expressed or distinguished within the current representational system. Additional search over that system explores more states but does not introduce the missing distinction. A representational or ontological revision is required. ■

This proposition marks an important difference between exhaustive search and conceptual discovery.

A search process asks:

Which object within this space satisfies the criterion?

A horizon-crossing process asks:

Is this the correct space, object type, or criterion?

The latter question belongs to metacognitive discovery.

Definition 12.31. Conceptual degree of freedom.
A conceptual degree of freedom is a revisable component of a discovery architecture, such as its representation, ontology, feature system, search grammar, objective, or validation standard.

Discovery is restricted when these degrees of freedom are fixed invisibly.

Definition 12.32. Reflexive search space.
A reflexive search space contains not only candidate solutions but also candidate representations, interpretations, objectives, and search procedures.

A conventional search has the form

search over objects.

A reflexive search has the form

search over objects and ways of searching for objects.

The second is vastly larger and requires stronger correction, provenance, and stopping rules.

Definition 12.33. Epistemic membrane.
An epistemic membrane is the controlled boundary between speculative generation and validated formal content.

The membrane must be permeable enough to permit formal return, but restrictive enough to prevent unsupported promotion.

Proposition 12.34. Discovery requires both permeability and separation.
A completely impermeable boundary between interpretation and formalization prevents creative hypotheses from entering mathematics, while the absence of a boundary permits speculation to masquerade as established result.

Proof.
If no interpretive claim can be formalized, abductive generation cannot contribute to mathematical development. If every interpretive claim is treated as formal, validation status is lost. Productive discovery therefore requires a controlled transition between distinct domains.
■

Definition 12.35. Epistemic choreography.
Epistemic choreography is the ordered coordination of expansion, selection, correction, formalization, and validation across one or more agents.

The term emphasizes that successful discovery depends not only on which operations are available but also on when they are applied.

Too-early pruning suppresses alternatives.

Too-late pruning produces saturation.

Too-early formalization freezes arbitrary choices.

Too-late formalization leaves speculation untestable.

Too-little reflection repeats failure.

Too-much reflection prevents return.

Discovery architecture is therefore partly a problem of timing.

Definition 12.36. Discovery rhythm.
Discovery rhythm is the temporal pattern by which a process alternates among expansion, concentration, testing, correction, reflection, and formal consolidation.

A productive rhythm may be represented schematically as

diverge → compare → test → correct → converge → formalize → reopen.

The final reopening is significant. Validation closes a particular formal obligation but need not terminate interpretation. A proved theorem may create a new presented form and initiate another cycle.

Thus discovery is recursively generative:

validated result → new observation.

Theorem 12.37. Recursive discovery principle.
Every validated result may function as a new presented form and thereby initiate a further discovery process.

Proof.
A validated result possesses structure that may be represented, compared, interpreted, generalized, restricted, translated, or placed in tension with other results. It can therefore occupy the observational role in a new discovery state. ■

The endpoint of one discovery is consequently the beginning of another.

The proposed vocabulary does not constitute a completed science of discovery. Several terms will require refinement, operationalization, historical testing, and comparison with established theories. Some distinctions may prove too coarse; others may divide phenomena that are better treated together.

Nevertheless, naming these operations has methodological value. What remains unnamed is difficult to record. What is not recorded is difficult to compare. What cannot be compared is difficult to improve.

The governing thesis of this paper is therefore:

Mathematical discovery becomes more intelligible when its interpretive branches, productive failures, corrections, reflective transitions, and provenance are treated as formal objects of study.

Proof remains indispensable. It is the principal mechanism by which mathematical claims acquire justified status. But proof is one region of a larger epistemic architecture.

A complete account of mathematics must therefore distinguish:

what is true;

why it is true;

how it was found;

how its discovery process was corrected;

and how that history can generate further mathematics.

13. References

Aliseda, A. *Abductive Reasoning: Logical Investigations into Discovery and Explanation*. Springer, 2006.

Boden, M. A. *The Creative Mind: Myths and Mechanisms*. 2nd ed., Routledge, 2004.

Cellucci, C. *Rethinking Logic: Logic in Relation to Mathematics, Evolution, and Method*. Springer, 2013.

Chalmers, A. F. *What Is This Thing Called Science?* 4th ed., Open University Press, 2013.

Darden, L. *Reasoning in Biological Discoveries: Essays on Mechanisms, Interfield Relations, and Anomaly Resolution*. Cambridge University Press, 2006.

Dunbar, K. "How Scientists Really Reason: Scientific Reasoning in Real-World Laboratories." In R. J. Sternberg and J. E. Davidson, eds., *The Nature of Insight*. MIT Press, 1995, pp. 365–395.

Feyerabend, P. *Against Method*. New Left Books, 1975.

Gabbay, D. M., and Woods, J. *The Reach of Abduction: Insight and Trial*. Elsevier, 2005.

Gentner, D. "Structure-Mapping: A Theoretical Framework for Analogy." *Cognitive Science* 7, 1983, pp. 155–170.

Giere, R. N. *Explaining Science: A Cognitive Approach*. University of Chicago Press, 1988.

Hadamard, J. *The Psychology of Invention in the Mathematical Field*. Princeton University Press, 1945.

Hanson, N. R. *Patterns of Discovery: An Inquiry into the Conceptual Foundations of Science*. Cambridge University Press, 1958.

Hintikka, J. *Socratic Epistemology: Explorations of Knowledge-Seeking by Questioning*. Cambridge University Press, 2007.

Holyoak, K. J., and Thagard, P. *Mental Leaps: Analogy in Creative Thought*. MIT Press, 1995.

Kitcher, P. *The Nature of Mathematical Knowledge*. Oxford University Press, 1984.

Kuhn, T. S. *The Structure of Scientific Revolutions*. 2nd ed., University of Chicago Press, 1970.

Lakatos, I. *Proofs and Refutations: The Logic of Mathematical Discovery*. Cambridge University Press, 1976.

Langley, P., Simon, H. A., Bradshaw, G. L., and Zytkow, J. M. *Scientific Discovery: Computational Explorations of the Creative Processes*. MIT Press, 1987.

Magnani, L. *Abduction, Reason, and Science: Processes of Discovery and Explanation*. Kluwer Academic/Plenum Publishers, 2001.

Magnani, L. *Abductive Cognition: The Epistemological and Eco-Cognitive Dimensions of Hypothetical Reasoning*. Springer, 2009.

Mancosu, P., Jørgensen, K. F., and Pedersen, S. A., eds. *Visualization, Explanation and Reasoning Styles in Mathematics*. Springer, 2005.

Newell, A., and Simon, H. A. *Human Problem Solving*. Prentice-Hall, 1972.

Nickles, T., ed. *Scientific Discovery, Logic, and Rationality*. D. Reidel, 1980.

Peirce, C. S. *Collected Papers of Charles Sanders Peirce*. Vols. 1–8. Harvard University Press, 1931–1958.

Poincaré, H. *Science and Method*. Thomas Nelson and Sons, 1914.

Polya, G. *How to Solve It*. Princeton University Press, 1945.

Polya, G. *Mathematics and Plausible Reasoning*. 2 vols., Princeton University Press, 1954.

Popper, K. R. *The Logic of Scientific Discovery*. Hutchinson, 1959.

Schön, D. A. *The Reflective Practitioner: How Professionals Think in Action*. Basic Books, 1983.

Simon, H. A. *The Sciences of the Artificial*. 3rd ed., MIT Press, 1996.

Thagard, P. *Conceptual Revolutions*. Princeton University Press, 1992.

Toulmin, S. *The Uses of Argument*. Cambridge University Press, 1958.

Wallas, G. *The Art of Thought*. Jonathan Cape, 1926.

Wimsatt, W. C. *Re-Engineering Philosophy for Limited Beings: Piecewise Approximations to Reality*. Harvard University Press, 2007.